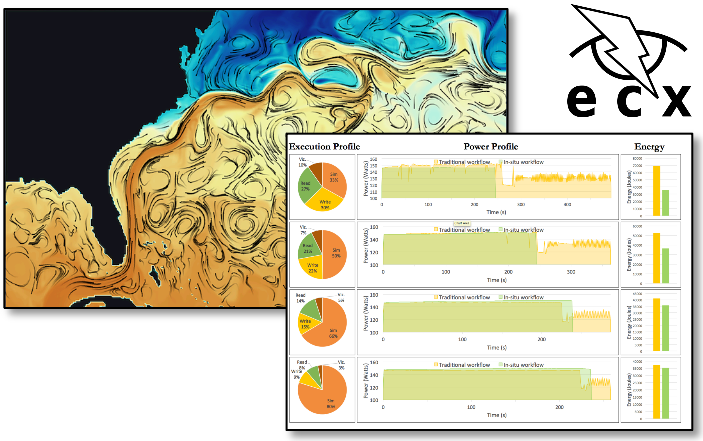
Optimizing The Energy Usage and Cognitive Value of Extreme Scale Data Analysis Approaches

Contact: James Ahrens, ahrens@lanl.gov (LA-UR-15-20105)



The Challenge of Extreme Scale

Scientific discovery at the extreme scale is a unique technical challenge, requiring the reduction of massive amounts of data into compact analysis products that capture key scientific insights. This analysis process needs to occur under extreme scale computational constraints including minimizing *data movement*, *energy usage* and *storage usage*. Put simply, extreme scale computing platforms are to achieve a three orders of magnitude increase in computational performance while consuming only two times the electrical power of current platforms. Data movement costs will dominate energy usage at this scale, so the HPC community expects extreme scale analysis algorithms will be utilized to reduce simulation results *in situ* — during the simulation run. This reduction will occur, broadly speaking, via some type of adaptive sampling, such as signal, statistical or feature-based sampling.

We consider the computational platform and the scientists using the platform as a system to be optimized. Our goal is to maximize scientific insight from sampled results while minimizing power.

We propose to use the unifying structure of a scientific data analysis workflow to knit together the research efforts of experts in *extreme data analysis, supercomputing energy usage, and scalable algorithms* with experts in *perception and human cognition* to explore methods for optimizing the combined computation/cognition system we will encounter at extreme scale. Continuing our very successful and longstanding collaborations, we will work closely with ASCR scientific simulation partners including climatologists, cosmologists and plasma physicists to design, evaluate and deploy our work.

We propose to investigate how changes in our sampling algorithms—necessary because of exascale power constraints—impact the cognitive value of the resulting data. Our goal is perceptual and cognitive optimization of tools that enhance analytic workflows while minimizing power consumption. We will pursue a three phase approach to the research, isolating specific portions of the potentially vast experimental space, in order to be able to draw conclusions across scales—from supercomputers to compute nodes, down to the subsystems of the compute node. We have assembled a team of experts from the diverse areas of Algorithms, Power, and Cognition/Perception to address the issues inherent in the optimization of the human/compute system at exascale.

Goal Our goal is to significantly improve the way analysis algorithms are developed, measured and used at the petascale and in the future at the exascale. The project will move beyond simple performance measurements to assess the value of analysis algorithms. Our goal is perceptual optimization of visualization algorithms and cognitive optimization of interactive methods in visualization tools.

Experimental Design Methodology

As a guiding framework for our proposal we will use an experiment design process to organize our work. Our expectation is that by observing our workflows and systematically experimenting with different input parameter settings we will understand their impact on the resulting power usage and analysis output quality. The input parameters we are considering are:

1) Data sampling approaches
2) How these are configured to run on a computational infrastructure and
3) Visual presentation approaches of the sampled data.

More formally, an *experiment* incorporates input parameters and produces output response variables to understand cause and effect relationships. We refer to *statistical experimental design* (SED) as the use of statistical techniques to vary input parameters as efficiently and effectively as possible in order to derive their relationships to output responses. In summary, our *goal* is to understand the cause and effect of sampling approaches and their mapping to a computational infrastructure on both power usage and cognitive/scientific understanding. A significant benefit of using an SED approach is that it provides a framework for our diverse set of algorithmic, power and cognition researchers to align and focus research.

Our Team: Experts in Energy, Algorithms and Human Perception/Cognition

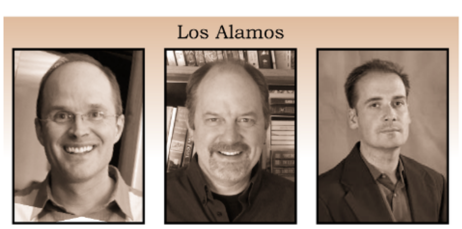
We have assembled a team of experts in three important areas–Energy, Algorithms and Human Perception/Cognition–who will work together to understand and model the interaction among power, sampling approaches, and visual representations.

The work will occur in three phases:

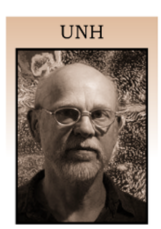
1) **Establishment of End-to-End Workflow measurements.** Initial power measurements, identification of essential cognitive and perceptual tasks in the target science domains, and initial experiments measuring effectiveness of sample visualizations
2) **Measuring the Behavior of Sampling-Modified Workflow.** The team will measure how specific sampling algorithms impact the power requirements of analysis and the visualization algorithms used to interact with the data.
3) **Modeling and Optimization of a Power-constrained Workflow.** Using experimental results from Phase 1 and 2, we will create a surrogate model that captures the relationships between our input parameters (e.g. data sampling approaches, configurations of workflows on computational infrastructures, and visual representations of the sampled data) and output parameters (e.g. power usage and cognitive value). This will then be used to answer optimization and what-if scenario questions posed by scientists.

Early Successes and Impacts

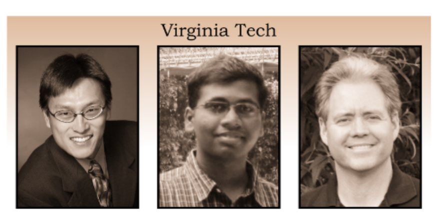
A paper, "Colormaps That Improve Perception of High-Resolution Ocean Data," has been accepted to CHI 2015. An initial perception experiment has been designed and is being run. In addition, the team has performed early measurements of power for representative *in situ*



Jim Ahrens David Rogers Jon Woodring



Colin Ware



Wu Feng Vignesh Adhinarayanan Mark Gardner

Francesca Samsel Terry Geld Greg Abram